THE PROTEIN SOCIETY

ACCELERATED COMMUNICATIONS

Structural propensities of kinase family proteins from a Potts model of residue co-variation

Allan Haldane,[1] William F. Flynn,[1,2] Peng He,[1] R.S.K. Vijayan,[1] and Ronald M. Levy[1]*

[1]Department of Chemistry, Center for Biophysics and Computational Biology, Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122

[2]Department of Physics and Astronomy, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854

Received 7 April 2016; Accepted 26 May 2016
DOI: 10.1002/pro.2954
Published online 31 May 2016 proteinscience.org

Abstract: Understanding the conformational propensities of proteins is key to solving many problems in structural biology and biophysics. The co-variation of pairs of mutations contained in multiple sequence alignments of protein families can be used to build a Potts Hamiltonian model of the sequence patterns which accurately predicts structural contacts. This observation paves the way to develop deeper connections between evolutionary fitness landscapes of entire protein families and the corresponding free energy landscapes which determine the conformational propensities of individual proteins. Using statistical energies determined from the Potts model and an alignment of 2896 PDB structures, we predict the propensity for particular kinase family proteins to assume a "DFG-out" conformation implicated in the susceptibility of some kinases to type-II inhibitors, and validate the predictions by comparison with the observed structural propensities of the corresponding proteins and experimental binding affinity data. We decompose the statistical energies to investigate which interactions contribute the most to the conformational preference for particular sequences and the corresponding proteins. We find that interactions involving the activation loop and the C-helix and HRD motif are primarily responsible for stabilizing the DFG-in state. This work illustrates how structural free energy landscapes and fitness landscapes of proteins can be used in an integrated way, and in the context of kinase family proteins, can potentially impact therapeutic design strategies.

Keywords: co-evolution; kinase; Potts model; inhibitor specificity; conformational preference

Abbreviations: DCA, direct coupling analysis; MSA, multiple sequence alignment

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH; Grant number: GM30580.

*Correspondence to: Ronald M. Levy, Center for Biophysics and Computational Biology, Department of Chemistry, Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122. E-mail: ronlevy@temple.edu

Introduction

A protein's sequence determines its free energy landscape, but it has proven a major challenge to predict sequence-dependent structural propensities from physical first principles. This has important practical consequences for therapeutic design, as conformational preferences can determine drug specificity. The type-II kinase inhibitor Gleevec is a prime

example, as it binds strongly to ABL kinase yet not to SRC kinase despite their having 47% sequence identity.[1,2] Gleevec's specificity has been suggested to be due in part to differing propensities of kinase proteins for a conformation known as "DFG-out" which the protein must take on in order to bind type-II inhibitors.[3–7] However this has been disputed and the sequence-dependent origins of the difference have proven difficult to confirm purely through structural analysis.[8–11]

The evolutionary origins of proteins open another angle of attack. Physical interactions between two residues in a protein's structure leads to their mutational co-variation in a multiple sequence alignment (MSA) of the protein family, which has motivated coevolutionary analysis techniques which predict contacts in structure by identifying strongly correlated position-pairs in the MSA (see Refs. 12, 13 for review). "Inverse Ising" methods have proven particularly suited for this purpose. These infer a statistical-energetic "Potts" Hamiltonian model whose parameters correspond to direct pairwise residue–residue interaction strengths, by fitting MSA statistics using techniques borrowed from statistical physics.[14] The power of inverse Ising inference has been demonstrated through its use as the central component of "direct coupling analysis" (DCA) for protein contact prediction, which has been shown to predict the top 200 intra-protein contact pairs in many proteins with ~80% accuracy as confirmed by X-Ray Crystallography and NMR studies, as well as inter-protein contacts, alternative uncrystallized conformations, ligand-mediated contacts, and it has been used for ab-initio structure prediction.[14–20]

The Potts model can be used for more than predicting contacts. The model provides a probability (or with a logarithm, a statistical energy) of any given sequence, and predicts the change in a sequence's statistical energy for any set of mutations. This statistical energy is related to the folding free energy of the protein, and can be decomposed into position- and residue-specific interaction terms whose relationship with the pairwise terms in structure-based free energy functions is just beginning to be explored.[21,22] This raises the possibility of predicting sequence-specific properties including conformational propensities.

Our goal is to infer conformational propensities of individual kinases for the inactive DFG-out state, in which a "DFG" motif is oriented away from the kinase's active site unlike in the active DFG-in conformation.[6] We predict conformational preference by "threading" calculations of the Potts energy for a sequence as a function of conformation. Sequences predicted by our analysis to have a high penalty for the DFG-out state are never observed in that state in crystal structures, while the remaining sequences are observed in both DFG-in and DFG-out, and we also find that sequences with high predicted penalty bind poorly to type-II inhibitors in a high-throughput binding assay. Furthermore, our analysis suggests that the stability of the activation loop in the DFG-in state plays an important role in contouring the energy landscape.

Inference Based on Correlated Sequence Variation

Inverse Ising methods infer a statistical model $P(S)$ for the probability of observation of a sequence S which captures the statistical features of a MSA of a protein family up to second order, in the form of the univariate and bivariate marginals (frequencies) f^i_α and $f^{ij}_{\alpha\beta}$ of the residues at each position and each position-pair i, j, for residue identities α, β. The maximum entropy (least biased) model which reproduces the observed bivariate marginals takes the form $P(S) \propto e^{-E(S)}$ where $E(S)$ is the statistical energy, given by the Potts Hamiltonian $E(S)=\sum_i^L h^i_{S_i}+\sum_{i<j}^L J^{ij}_{S_iS_j}$ where the model parameters $h^i_{S_i}$ ("fields") represent the statistical energy of residue S_i at position i, and $J^{ij}_{S_iS_j}$ ("couplings") represent the energy contribution of a position-pair i, j. This model has been of interest in protein structure prediction because strong couplings $J^{ij}_{\alpha\beta}$ are expected to correspond to direct physical interactions in the protein 3d structure, in contrast to the evolutionary correlations $C^{ij}_{\alpha\beta}=f^{ij}_{\alpha\beta}-f^i_\alpha f^j_\beta$ which reflect both direct and indirect interactions.[14,18]

Determining the values of Potts couplings given bivariate marginals is a significant computational challenge known as the inverse Ising problem, and a variety of algorithms have been devised to solve it.[15,18,23–31] We have elaborated on a quasi-Newton Monte Carlo method[32,33] which is more computationally intensive but yields a more accurate model, and adapted it for protein family coevolutionary analysis with a highly parallel implementation for GPUs. To reduce the size of the problem and reduce the effect of sampling error, we use a reduced amino acid alphabet of 8 characters, chosen independently at each position in a way which preserves the correlation structure of the MSA (see methods).

Extracting Conformational Information from the Potts Model and Crystal Structures

In typical applications of DCA an overall "interaction" score is calculated for each position-pair based on the coupling parameters and a threshold determines predicted interactions, which have been used to bias coarse grained molecular simulations.[19,31] Contact prediction is illustrated in Figure 1A (upper triangle), where the 64 coupling values for each position-pair are summarized using a weighted "Frobenius norm" (described in SI text) into a single number, shown

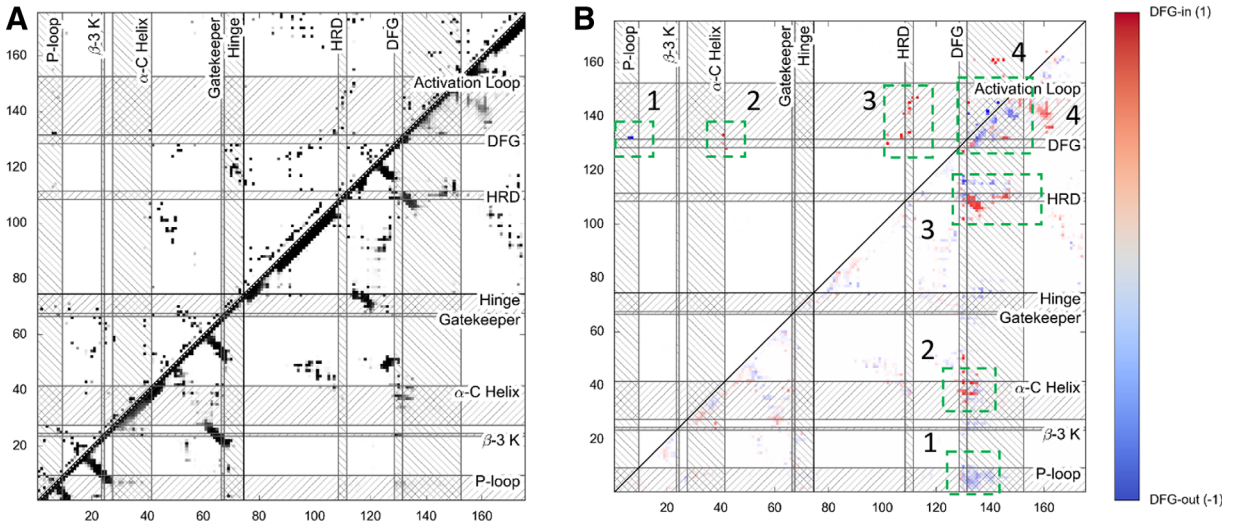

Figure 1. Contact prediction using the Potts model. (A) Potts model predicted contacts computed using the weighted Frobenius Norm (upper triangle), and a heatmap of crystal structure contact frequency at 6Å cutoff for each residue pair (lower triangle). Important structural motifs such as the DFG and HRD triplets are annotated as hashed rows and columns. (B) Difference in contact frequency in the DFG-in and DFG-out conformations, based on PDB structures (lower triangle), with corresponding high-Frobenius-Norm pairs highlighted in matching colors (upper triangle). The contact frequency was computed separately for the DFG-out and DFG-in structures and subtracted, giving a value from −1 to 1.

as a heatmap. We also align 2896 kinase PDB structures and count the frequency of residue–residue contacts with a 6Å distance cutoff, shown as a complementary heatmap (lower triangle, Fig. 1A). The correspondence between the two maps is striking, demonstrating how the Potts model contains information about specific interactions within the protein.

In Figure 1B, lower triangle, we show the difference in contact frequency between the DFG-in and DFG-out conformations based on a PDB crystal structure classification (see methods). Contacts shared by both conformations corresponding to the overall fold cancel out, highlighting position-pairs which differentiate the conformations. The Potts model predicts strong coevolutionary interactions at many of these positions (upper triangle) suggesting it may be used to understand the conformational transition.

In particular, this analysis highlights the importance of the activation loop in the conformational transition and identifies specific interactions it takes part in. Figure 1B shows four relevant regions whose structures are illustrated in Figure 2. Interactions in region 1 between the activation loop and the P-loop are much more common in the DFG-out state as has been previously reported,[6,36,37] and the co-evolutionary analysis predicts two strongly interacting pairs, (6,132) and (7,132), where 132 is the DFG + 1 position (see numbering in Supporting Information table S2). In region 2, residues near the DFG motif interact with the C-helix in the DFG-in state,[36,38] as a result of a network known as the R-spine which is broken in DFG-out.[6,37,39] The Potts model predicts a strong interaction between the DFG + 2 and DFG-1 residues and the end of the C-helix. Region 3 corresponds to interaction between the HRD motif and activation loop. In DFG-in this loop forms a beta-strand near the C-helix, while in DFG-out it folds to form a more distant two-turn helix.[6,37,39] The R of the HRD motif is in contact with this beta-strand in DFG-in, and the Potts model predicts a number of interactions in this region. Region 4 illustrates self-contacts in the activation loop in DFG-out due to its more folded and compact conformation. All four regions illustrate major global differences between DFG-in and DFG-out reported in recent publications.

Predicting Kinase Sequence Conformational Preference

The Potts energy can be decomposed into position- and residue- specific components (the fields and couplings), allowing analysis of the statistical energy of regions within a sequence and the energetic coupling between particular position pairs of that sequence. By evaluating the Potts statistical energy for a given sequence (not necessarily from the PDB dataset) only including coupling terms corresponding to positions observed to be in contact in a chosen structure (defined by a 6Å atom-atom cutoff distance), we obtain a "threaded" energy of the sequence in that conformation. A single sequence may be evaluated over multiple conformations.

We use this threading calculation to model the DFG-in to DFG-out transition. We order the 2869 kinase PDB structures by DFG-in and DFG-out conformation using PCA analysis (see methods). We find the second principal component distinguishes DFG-in (PCA2 < 40) and DFG-out (PCA2 > 48)

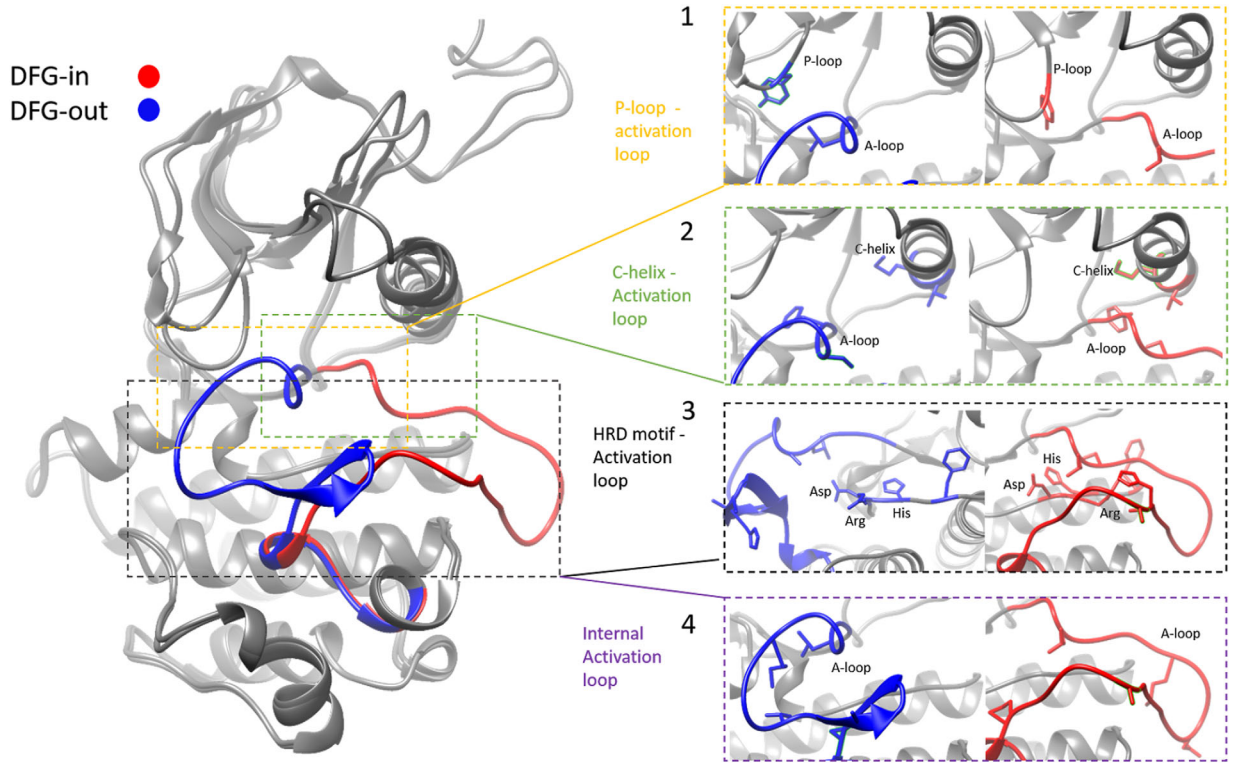

Figure 2. Structural differences between the DFG-in and DFG-out conformations of ABL kinase. Left: DFG-in (red, 2GQG [34]) and DFG-out (blue, 1IEP [35]) structures superimposed, showing the activation loop extended in the DFG-in state. Right: Highlighted regions from figure 1B showing DFG-out (left) and DFG-in (right), with Potts predicted interacting residues shown as sticks. Region 1: P-loop to activation loop interactions, showing Y7 and L132 (DFG + 1). Region 2: C-helix to activation loop interactions, showing V41, M42, F130,and S133. Region 3: HRD motif to activation loop interaction, showing F107, HRD (109–111), A143, H144, and A145. Region 4: Interactions internal to the activation loop, showing L132, M136, T140, A143, and P147.

according to the KLIFS database. By averaging a sequence's threaded energy over this conformational parameter we obtain an effective "potential of mean force" for that sequence [Fig. 3(A)], showing that certain sequences have a greater relative penalty for taking the DFG-out conformation. Recent work suggests it is possible to connect the Potts statistical energy to physical units; we estimate that the scale shown in Figure 3 to be 2–3 kcal/mol based on the analysis in Ref. 21.

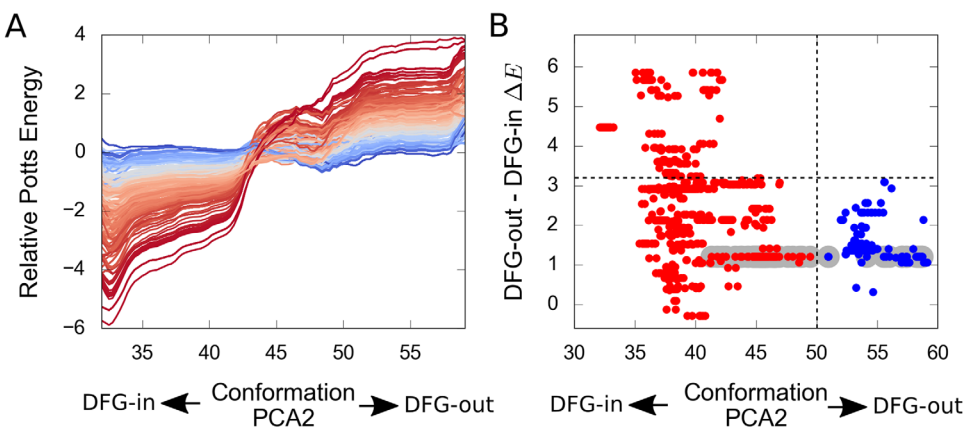

Figure 3. Prediction of conformational penalties. (A) Predicted conformational energy of PDB sequences as a function of conformation (averaged over 5-unit windows along the conformational axis, with mean value subtracted), colored according the DFG-out penalty score (blue is low penalty). (B) Predicted DFG-out penalty versus conformation for each sequence and structure in the KLIFS database. The dashed horizontal line distinguishes sequences with a low DFG-out penalty from those with high penalty, and the vertical line separates DFG-in (red) and DFG-out (blue) structures. The PDB structures of p38a MAP kinase are shadowed in gray.

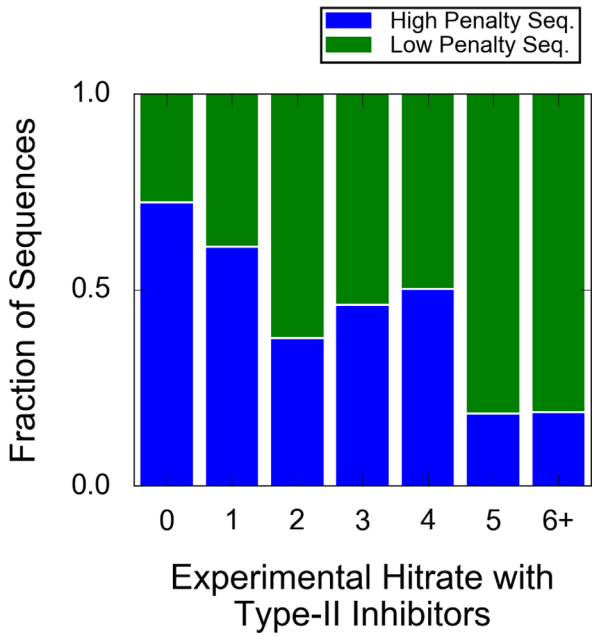

Figure 4. Comparison of predicted DFG-out penalty to measured type-II inhibitor hit-rate, determined from a binding assay against 13 type-II inhibitors. The fraction of low versus high penalty sequences is plotted for each hit-rate.

We compute an overall DFG-out "penalty" score for an arbitrary kinase sequence as the difference in its average threaded energy between DFG-in and DFG-out conformations. This is equivalent to computing $\Delta E(s)=\sum_{i<j} J^{i,j}_{s_i,s_j}(c^{\text{out}}_{i,j}-c^{\text{in}}_{i,j})$ where $c^{\text{in}}_{i,j}$ and $c^{\text{out}}_{i,j}$ are the contact frequency of the pair i,j in the two conformational states, reflecting a penalty for the sequence to take on the DFG-out conformation.

Validating this predicted penalty score using sequences from the PDB, we find that only sequences with low predicted DFG-out penalty are observed in the DFG-out conformation in the KLIFS structural annotation [Fig. 3(B)]. Many sequences with low DFG-out penalty are observed in both the DFG-in and DFG-out conformations, for example p38a MAP kinase [Fig. 3(B), gray], which is expected since the active DFG-in state is necessary for kinase function.

We also compare our predictions to experimental results from a high-throughput inhibitor binding assay of 299 human kinases sequences against 13 type-II inhibitors[40] (see methods). The number of inhibitors which bind to a kinase out of the 13 (its "hit-rate") is an experimental measure of the kinase's conformational penalty, which averages out the effects of ligand-specific interactions. In Figure 4 we show that sequences with low hit-rate are predominantly predicted to have high penalty (>3) for the DFG-out state necessary for type-II inhibitor binding. The difference in mean hit-rate of 2.1 between the high and low penalty sequences is highly significant ($P<10^{-10}$), as measured by a permutation test. This result supports the role of conformational preferences in determining drug susceptibility suggested by computational studies.[2,3,41]

The highest penalty kinase in this analysis, Aurora A kinase, has no type-II inhibitors developed or reported in development and does not have typical DFG-out structures in the PDB.[6] Aurora A has a conformation called DFG-up which has some similarities to DFG-out but is different enough to be classified separately.[42,43] There are reported structures in which the DFG motif is 'out' (2C6E, 2J4Z,[44] not in the KLIFS database) but the activation loop is in the DFG-in-like extended form, and they are not bound to type-II inhibitors. The DFG motif in these unusual structures may be forced to an out-like state by ligand-specific interactions. This further suggests the activation loop itself contributes to the conformational preference.

Stabilization of the Activation Loop in DFG-in by Particular Position Pairs

We examine the highest DFG-out penalty sequences to determine which interactions contribute to their high penalty. These positions are among those highlighted in the green boxes in Figure 2. We find that these positions have significantly more favorable couplings in DFG-in sequences than in DFG-out sequences on average, and from structural analysis these position pairs make frequent contacts within 6Å in the DFG-in state but not in DFG-out state. This suggests stabilization of the activation loop in the DFG-in sequences, which is consistent with a recent computational study which showed that certain mutants stabilize the activation loop, biasing the protein towards the DFG-in state.[41]

Conclusions

Free energy calculations have confirmed that single point mutations which lead to resistance to type-II inhibitors act by increasing the DFG-in to DFG-out conformational free energy penalty.[2,6,41] We have shown that by examining the position- and sequence- specific components of the Potts Hamiltonian it is possible to predict this sequence-dependent penalty, deepening the link between the evolutionary fitness and energy landscapes of proteins, which we hope will have an impact on therapeutic design strategies.

Methods

Sequence datasets

We use HHblits[45] to search the Uniprot database with the Pfam kinase family seed (PF00069), obtaining 127113 kinase sequences after filtering for valid kinases. We weight these sequences to account for phylogenetic and experimental biases at a 40%

identity threshold using the weighting strategy in Ref. 18, leaving $N = 8149$ effective sequences with 175 positions. To reduce the alphabet size we randomly merge pairs of letters at each position which, when treated as identical, would minimize the root mean square difference between the Mutual Information (MI) scores for all position pairs in the reduced alphabet and full 21 letter alphabet, until all positions have been reduced to 8 letters.

Potts model inference

For a set of trial couplings we estimate bivariate marginals by Markov Chain Monte Carlo (MCMC) evolution of 131072 sequences in parallel according to the Potts Hamiltonian, each for 6.4 million steps to reach equilibrium. The residuals relative to the dataset marginals are then used in a quasi-Newton update step of the Hamiltonian parameters (see Supporting Information text for details).

PDB structure datasets

We obtain 2869 kinase crystal structures from the PDB and align them to our sequence dataset. We choose 351 atom–atom pairs which may be related to the DFG-in to DFG-out transition whose distances we use as variables for PCA analysis (see Supporting Information Fig. S2). After filtering based on the PCA analysis, we find 432 structures annotated as DFG-in and 93 as DFG-out in the KLIFS database.[46] Contacts are computed based on closest atom-atom distances. When averaging, sequences are weighted using a 40% identity threshold, renormalizing to account for unresolved residues.

Kinase binding assay

We filter the 442 kinases tested in a binding assay[40] keeping only the 299 unique non-mutant kinases with a complete catalytic domain sequence in the Entrez Genbank database,[47] and predict their penalty scores. An inhibitor "hit" was counted for any assay with dissociation constant $< 10\ \mu M$.

References

1. Lovera S, Sutto L, Boubeva R, Scapozza L, Dlker N, Gervasio FL (2012) The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation. J Am Chem Soc 134:2496–2499.
2. Lin YL, Meng Y, Jiang W, Roux B (2013) Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. Proc Natl Acad Sci 110:1664–1669.
3. Aleksandrov A, Simonson T (2010) Molecular dynamics simulations show that conformational selection governs the binding preferences of Imatinib for several tyrosine kinases. J Biol Chem 285:13807–13815.
4. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J (2000) Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. Science 289:1938–1942.
5. Seeliger MA, Nagar B, Frank F, Cao X, Henderson MN, Kuriyan J (2007) c-Src binds to the cancer drug Imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. Structure 15:299–311.
6. Vijayan RSK, He P, Modi V, Duong-Ly KC, Ma H, Peterson JR, Dunbrack RL, Levy RM (2015) Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. J Med Chem 58:466–479.
7. Liu Y, Gray NS (2006) Rational design of inhibitors that bind to inactive kinase conformations. Nat Chem Biol 2:358–364.
8. Mbitz H (2015) The ABC of protein kinase conformations. Biochim et Biophys Acta 1854:1555–1566.
9. Seeliger MA, Ranjitkar P, Kasap C, Shan Y, Shaw DE, Shah NP, Kuriyan J, Maly DJ (2009) Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations. Cancer Res 69: 2384–2392.
10. Wilson C, Agafonov RV, Hoemberger M, Kutter S, Zorba A, Halpin J, Buosi V, Otten R, Waterman D, Theobald DL, Kern D (2015) Using ancient protein kinases to unravel a modern cancer drugs mechanism. Science 347:882–886.
11. Agafonov RV, Wilson C, Otten R, Buosi V, Kern D (2014) Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. Nat Struct Mol Biol 21:848–853.
12. de Juan D, Pazos F, Valencia A (2013) Emerging methods in protein co-evolution. Nat Rev Genet 14:249–261.
13. Marks DS, Hopf TA, Sander C (2012) Protein structure prediction from sequence variation. Nat Biotech 30: 1072–1080.
14. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T (2009) Identification of direct residue contacts in proteinprotein interaction by message passing. pnas 106: 67–72.
15. Lvkvist C, Lan Y, Weigt M, Aurell E, Ekeberg M (2013) Improved contact prediction in proteins: using pseudolikelihoods to infer potts models. Phys Rev E 87: 012707.
16. Morcos F, Jana B, Hwa T, Onuchic JN (2013) Coevolutionary signals across protein lineages help capture multiple protein conformations. Proc Natl Acad Sci 110:20533–20538.
17. Sukowska JI, Morcos F, Weigt M, Hwa T, Onuchic JN (2012) Genomics-aided structure prediction. Proc Natl Acad Sci 109:10340–10345.
18. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci 108:E1293–E1301.
19. dos Santos RN, Morcos F, Jana B, Andricopulo AD, Onuchic JN (2015) Dimeric interactions and complex formation using direct coevolutionary couplings. Sci Rep 5:13652.
20. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C (2011) Protein 3D structure computed from evolutionary sequence variation. PLoS One 6:e28766EP.
21. Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG (2014) Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. Proc Natl Acad Sci 111:12408–12413.
22. Cheng RR, Raghunathan M, Noel JK, Onuchic JN (2015) Constructing sequence-dependent protein models using coevolutionary information. Prot Sci 25: 111122.

23. Lapedes A, Giraud B, Jarzynski C (2002) Using sequence alignments to predict protein structure and stability with high accuracy. arXiv:1207.2484
24. Mzard M, Mora T (2009) Constraint satisfaction problems and neural networks: a statistical physics perspective. J Physiol-Paris 103:107–113.
25. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T (2009) Identification of direct residue contacts in proteinprotein interaction by message passing. Proc Natl Acad Sci 106:67–72.
26. Monasson R, Cocco S (2011) Adaptive cluster expansion for inferring boltzmann machines with noisy data. Phys Rev Lett 106:090601.
27. Jones DT, Buchan DWA, Cozzetto D, Pontil M (2012) PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. Bioinformatics 28:184–190.
28. Balakrishnan S, Kamisetty H, Carbonell JG, Lee SI, Langmead CJ (2011) Learning generative models for protein fold families. Prot Struct, Funct, Bioinform 79: 1061–1078.
29. Burger L, van Nimwegen E (2010) Disentangling direct from indirect co-evolution of residues in protein alignments. PLoS Comput Biol 6:e1000633.
30. Baldassi C, Zamparo M, Feinauer C, Procaccini A, Zecchina R, Weigt M, Pagnani A (2014) Fast and accurate multivariate gaussian modeling of protein families: predicting residue contacts and protein-interaction partners. PLoS ONE 9:e92721EP.
31. Sutto L, Marsili S, Valencia A, Gervasio FL (2015) From residue coevolution to protein conformational ensembles and functional dynamics. Proc Natl Acad Sci 112:13567–13572.
32. Mora T, Bialek W (2011) Are biological systems poised at criticality? J Stat Phys 144:268–302.
33. Ferguson A, Mann J, Omarjee S, Ndungu T, Walker B, Chakraborty A (2013) Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. Immunity 38:606–617.
34. Tokarski JS, Newitt JA, Chang CYJ, Cheng JD, Wittekind M, Kiefer SE, Kish K, Lee FY, Borzillerri R, Lombardo LJ, Xie D, Zhang Y, Klei HE (2006) The structure of Dasatinib (BMS-0000) bound to activated ABL kinase domain elucidates its inhibitory activity against Imatinib-resistant ABL mutants. Cancer Research 66:5790–5797.
35. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, Clarkson B, Kuriyan J (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and Imatinib (STI–571). Cancer Research 62:4236–4243.
36. Guimares CRW, Rai BK, Munchhof MJ, Liu S, Wang J, Bhattacharya SK, Buckbinder L (2011) Understanding the impact of the P-loop conformation on kinase selectivity. J Chem Inf Model 51:1199–1204.
37. Taylor SS, Kornev AP (2011) Protein kinases: evolution of dynamic regulatory proteins. Trend Biochem Sci 36: 65–77.
38. Huang H, Zhao R, Dickson BM, Skeel RD, Post CB (2012) αC (alpha-C) helix as a switch in the conformational transition of Src/CDK-like kinase domains. J Phys Chem B 116:4465–4475.
39. Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE (2013) Transitions to catalytically inactive conformations in EGFR kinase. Proc Natl Acad Sci 110:7270–7275.
40. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP (2011) Comprehensive analysis of kinase inhibitor selectivity. Nat Biotech 29:1046–1051.
41. Lovera S, Morando M, Pucheta-Martinez E, Martinez-Torrecuadrada JL, Saladino G, Gervasio FL (2015) Towards a molecular understanding of the link between Imatinib resistance and kinase conformational dynamics. PLoS Comput Biol 11:e1004578EP.
42. Bavetsias V, Linardopoulos S (2015) Aurora kinase inhibitors: current status and outlook. Front Oncol 5:
43. Nikonova AS, Astsaturov I, Serebriiskii IG, Dunbrack RL, Golemis EA (2012) Aurora a kinase (AURKA) in normal and pathological cell division. Cell Mol Life Sci 70:661–687.
44. Heron NM, Anderson M, Blowers DP, Breed J, Eden JM, Green S, Hill GB, Johnson T, Jung FH, McMiken HH, Mortlock AA, Pannifer AD, Pauptit RA, Pink J, Roberts NJ, Rowsell S (2006) SAR and inhibitor complex structure determination of a novel class of potent and specific Aurora kinase inhibitors. Bioorg Med Chem Lett 16:1320–1323.
45. Remmert M, Biegert A, Hauser A, Soding J (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Meth 9:173–175.
46. van Linden OPJ, Kooistra AJ, Leurs R, de Esch IJP, de Graaf C (2014) KLIFS: a knowledge-based structural database to navigate kinaseligand interaction space. J Med Chem 57:249–277.
47. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) Genbank. Nucleic Acid Res 41:D36–D42.